SANDIA REPORT
SAND2000-2232
Unlimited Release
Printed November 2000

Development of a Discrete Spatial-Temporal SEIR Simulator for Modeling Infectious Diseases

Sean A. McKenna

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Prices available from (703) 605-6000
Web site: http://www.ntis.gov/ordering.htm

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A04
Microfiche copy: A01

SAND2000-2232
Unlimited Release
Printed November 2000

Development of a Discrete Spatial-Temporal SEIR Simulator for Modeling Infectious Diseases

Sean A. McKenna
Geohydrology Department
Sandia National Laboratories
PO Box 5800
Albuquerque, New Mexico 87185-0735

ABSTRACT

Multiple techniques have been developed to model the temporal evolution of infectious diseases. Some of these techniques have also been adapted to model the spatial evolution of the disease. This report examines the application of one such technique, the SEIR model, to the spatial and temporal evolution of disease. Applications of the SEIR model are reviewed briefly and an adaptation to the traditional SEIR model is presented. This adaptation allows for modeling the spatial evolution of the disease stages at the individual level. The transmission of the disease between individuals is modeled explicitly through the use of exposure likelihood functions rather than the global transmission rate applied to populations in the traditional implementation of the SEIR model. These adaptations allow for the consideration of spatially variable (heterogeneous) susceptibility and immunity within the population. The adaptations also allow for modeling both contagious and non-contagious diseases. The results of a number of numerical experiments to explore the effect of model parameters on the spread of an example disease are presented.

ACKNOWLEGEMENTS

This work has benefited from many fruitful discussions with Mark Koch, Al Zelicoff and Drayton Boozer. Careful reviews by Kevin Green and Erik Webb greatly improved the final content of this report. Funding for this work was provided by the LDRD program at Sandia through the efforts of Tom Karas.

Table of Contents

List of Figures

INTRODUCTION

The term "epidemic process" can be applied to a broad class of conceptual/numerical models that are used to characterize the spread of a condition through a population. Examples of epidemic processes include the spread of a contagious disease through a human population, the spread of a wildfire across a region of land, and the spread of a plant virus through an agricultural area. At the basis of all these applications is the transmission of the condition (e.g., disease, fire) from one object (individual or tree) to another, or to multiple objects (people or trees).

One class of models that has proven useful for describing epidemic processes is the SEIR model. The SEIR model describes four discrete states of an epidemic: Susceptiple, Exposed, Infectious and Recovered. Infectious individuals spread the disease to the susceptible population. Those in the susceptible population to which the disease is transmitted become exposed and after a period time, the incubation period, these individuals become infectious. Individuals remain infectious for a period of time, the infectious period, and then these individuals recover. In the majority of the SEIR model applications, the recovered individuals are considered to be immune to the disease for all remaining time.

The SEIR model has been used extensively to examine the temporal evolution of diseases. The origins of the SEIR model with applications to epidemiology go back to the Hamer-Soper model in the early 20th century (see: Hamer, 1906; Soper, 1929; and discussion in Grenfell, 1992). Modern applications of the SEIR model to epidemics began with the work of Bartlett in the 1950's (e.g., Bartlett, 1957). Much of the current work in the application of the SEIR model to understanding epidemics is focussed on determining the processes responsible for chaotic behavior in epidemics (e.g., Rohani, et al., 1999; Earn et al., 1999). A long-term data set (> 50 years) of the incidence of measles epidemics in England has been used for both development and verification of the SEIR model by a number of authors (e.g., Bartlett, 1956; Grenfell, 1992; Bolker and Grenfell, 1996). As pointed out by Bolker and Grenfell (1996), the continued failure of vaccination programs to eradicate measles in developed countries and the mortality rate among children due to measles in developing countries has sparked this long-term interest in modeling measles epidemics.

There are several motivations for this current study. The most urgent driver is the necessity to simulate syndromes, signs and symptoms, or "S^3" data at the individual patient level. These simulations will provide a hypothetical database against which pattern recognition algorithms can be tested to determine if the S^3 data and the accompanying spatial and temporal coordinates of the data can discriminate outbreaks of the flu or other common diseases from infectious diseases initiated by terrorist organizations. This same approach could also be used to detect the first occurrence of a disease in a location where it has not been previously recorded (e.g., the occurrence of West Nile Fever in New York City in 1999). Results of an SEIR model computed on a discrete grid of individuals will be used as the basis of the spatial and temporal evolution

of the S^3 data set. Eventually, actual S^3 data from the RSVP project at Sandia National Laboratories will be available as input to the pattern recognition portion of this project. However, it is doubtful that the RSVP program will be able to provide both background S^3 data and the S^3 data resulting from a bioterror attack. Therefore, it may still be necessary to simulate the S^3 data resulting from a bioterror attack even after the RSVP project becomes active.

The adaptations to the traditional SEIR completed in this study allow for the incorporation of spatial variability in the susceptibility/immunity of the population. Often, variability in these two population parameters is correlated with the ages of the individuals in the population and the age segregation of the population may display spatial variability. Therefore, a second motivation behind the construction of this model is to provide a tool for examining the effects of spatial variability in susceptibility and immunity within a population on the transmission of an infectious disease. Results of model runs completed with spatially variable parameters will be compared to results from the traditional SEIR model.

A third motivating factor in the development of an explicitly spatial SEIR model is the desire to examine geographically coordinated, or patterned, inoculation efforts on the spatial and temporal extent of a bio terror attack on a population. As an example, the concept of "ring vaccination", or vaccinating in a ring around the known areas of disease is generally credited with wiping out small pox in Africa.

TRADITIONAL SEIR MODEL

A number of authors have invoked the SEIR model to describe the temporal evolution of an infectious disease. There are subtle variations on the implementation of the SEIR model from one reference to the next. For consistency, the implementation used in Earn, et al. (1999) is followed here.

The SEIR model describes the progression of a disease through four stages: Susceptible, Exposed, Infectious and Recovered. This progression can be thought of as a simple, non-reversible, Markov chain where the state at time $t+1$ is solely a function of the state at time t. A set of difference equations models the progression of a disease through a population, or some fraction of a population.

The first stage is the susceptible stage, S, and the number of susceptible individuals in a population changes according to:

$$\frac{\Delta S}{\Delta t} = \nu N - \frac{(\beta I + \mu)S}{N}$$

where t is time [T, denotes the units of the variable with T being time], ν is the per capita birth rate [1/T], β is the rate of disease transmission between individuals [1/T], μ is the death rate [1/T] and I, S and N are the current number of infectious, susceptible and total

individuals in the population . In the traditional application of the SEIR model, β is the product of the inverse of the mean infectious period, γ [1/T], and the reproductive ratio of the infection, *Ro* [unitless]. The reproductive ratio is defined as the number of newly exposed individuals per infectious individual. The mean infectious period, $1/\gamma$ [T] is the average amount of time an infectious individual is able to expose susceptible individuals. The reproductive ratio is discussed further below.

The second stage of the disease is the exposed stage, *E*, with changes in this stage defined by:

$$\frac{\Delta E}{\Delta t} = \frac{\beta IS}{N} - (\sigma + \mu)E$$

where σ is the inverse of the mean incubation, period [1/T] and *E* is the number of exposed individuals in the population. The mean incubation period, $1/\sigma$ [T], is the average period of time an individual remains in the exposed, *E*, state. This time period from being exposed to being able to expose other susceptible individuals may include the time during which the patient becomes symptomatic.

The third stage of the disease is the infectious, *I*, stage. Changes to the infectious members of the population are defined as:

$$\frac{\Delta I}{\Delta t} = \sigma E - (\gamma + \mu)I$$

The final stage of the disease is the recovered stage, *R*:

$$\frac{\Delta R}{\Delta t} = \gamma I - \mu R$$

For the results presented in this report, it is assumed that once an individual reaches the R stage of the disease, that individual is completely immune to the disease. In the case of a fatal infectious disease, the members of this class are not recovered, but deceased.

It is noted by Earn et al. (1999) that if β is constant and Ro > 1, these equations lead to a stable equilibrium. This equilibrium may include the case where the disease spreads through the entire susceptible population and the equilibrium state is where all intially susceptible individuals have reached the recovered stage. If the β parameter varies seasonally, it is possible to get a broad range of dynamic behavior over multiple year time spans. Periodicity is often entered into the above equations through the use of a periodic disease transmission rate, $\beta(t)$. The periodicity is generally tied to the school year calendar with high contact rates during the school year and lower rates during summer vacation. It is noted here, that for the purposes of simulating infectious disease over the several months long time span of interest, the time span of a single bio terror attack, the above equations will not display chaotic behavior.

Simulations

The traditional SEIR model is capable of simulating the temporal evolution of an epidemic in a single population. As an example, the aggregate disease stage curves for the simulation of a contagious disease through a single population are shown in Figure 1. In this simulation, the population size is 100,000, the values of $\nu, \mu, 1/\sigma, 1/\gamma$ and Ro are set to 0.0, 0.0, 8.0 days, 5.0 days and 5.0 respectively. Setting the birth and death rates equal to zero ensures a constant population size. The initial fraction of exposed individuals is 0.001 and the length of the time step is set to 1.0 days. All of the simulations done in this report are deterministic- that is, all rates are single-valued and constant over time.

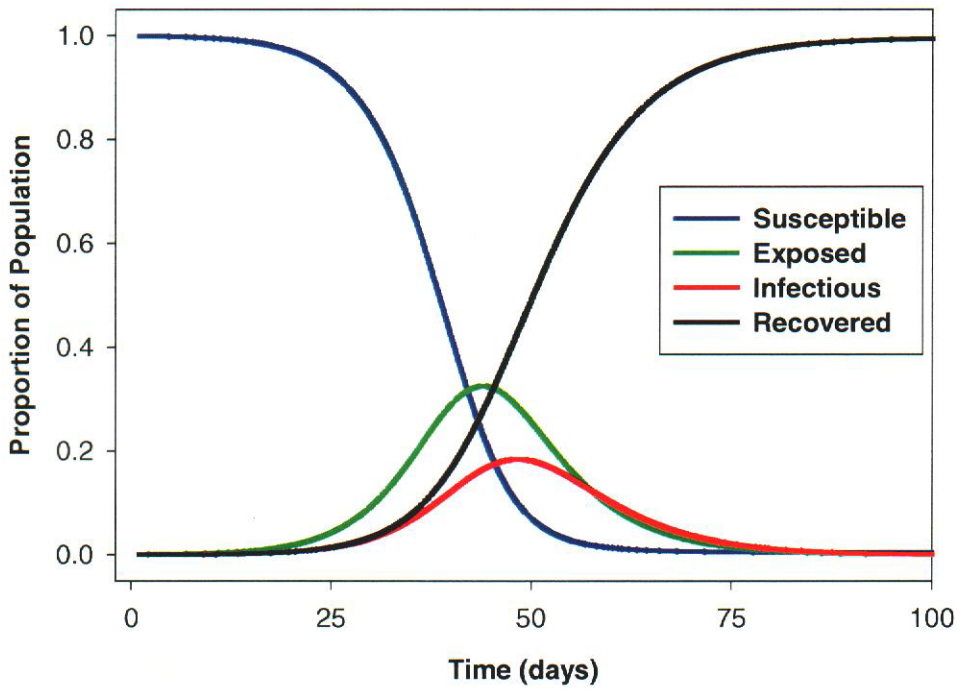

Figure 1. Example results from the SEIR model showing the temporal evolution of an epidemic through a population.

The results in Figure 1 show the evolution of the contagious disease through the entire population in roughly 100 days. Through the early stages of the epidemic and through the peak of the epidemic, the fraction of the population that is exposed is larger than the fraction that is infectious. This result is due to the incubation period being longer than the infectious period (8 days versus 5 days). For this report, the fraction of the population that is considered to be part of the epidemic at any one time is the sum of the exposed and infectious individuals. This sum is the number of individuals in the population that may be symptomatic and would thus seek out medical attention at any given time. In Figure 1, the fraction of the population in the epidemic is the sum of the Exposed and Infectious curves. The maximum proportion of the population that may be symptomatic occurs at 45 days when just less than one half the population (0.495) is exposed or infectious.

A critical parameter in the SEIR model is the reproductive ratio, *Ro*. This ratio is defined as the number of newly exposed individuals per infectious individual. If this ratio remains below 1.0, the disease cannot propagate through the population and the disease will run its course leaving a large fraction of the population still in the susceptible class. Values of *Ro* greater than 1.0 will spread the disease through greater and greater fractions of the population. It is noted that in the traditional SEIR model, *Ro* is real valued. The idea of one individual infecting a fraction of another individual does not have meaning in the real world or in a model where each individual is modeled discretely.

A set of simulations with the traditional SEIR model was done to examine the effect of *Ro* on the equilibrium values of the number of susceptible and recovered individuals in a population. The values of $1/\sigma$ and $1/\gamma$ are 8.0 and 5.0 days respectively. These values are deterministic (fixed) and represent the generally accepted values for measles (see Earn et al., 1999). The birth and death rates are set to zero to ensure a constant population size of 100,000 and the fraction of the population that is initially exposed is 0.001. The simulations are run as a closed system-no new exposed or infectious individuals are introduced to the system after the initial time step. The simulations are run out to 200 times the sum of $(1/\sigma + 1/\gamma)$ or 2600 days, and the time step is set to 1.0 days. This long time insures that the population reaches an equilibrium status. The results of this set of simulations are shown in Figure 2.

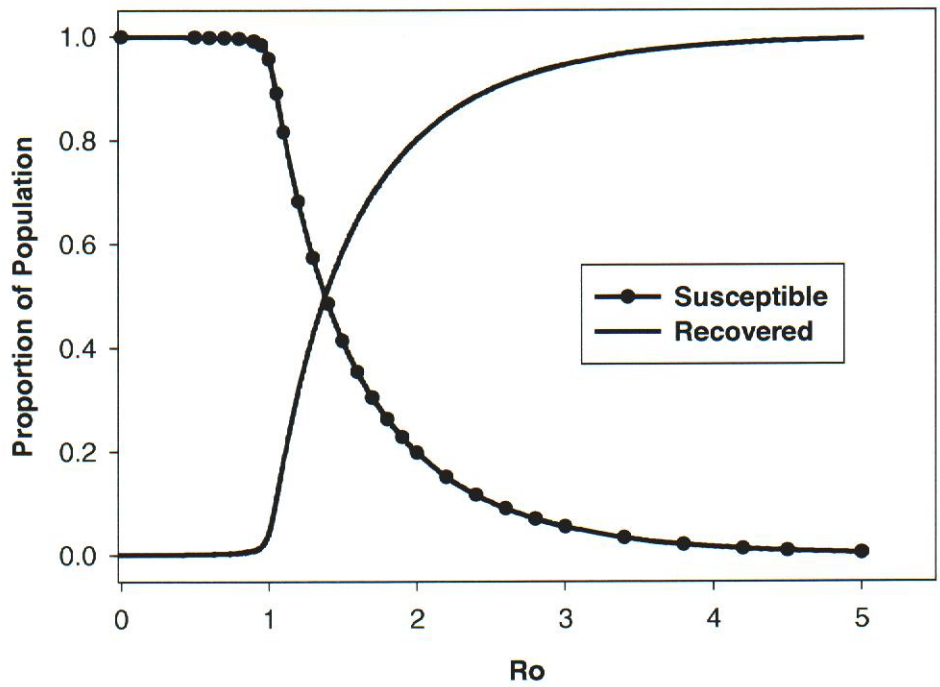

Figure 2. Proportion of Susceptible and Recovered individuals at equilibrium (2600 days) as a function of *Ro*.

The results of these simulations demonstrate the threshold behavior of the population as a function of *Ro*. As *Ro* increases beyond 1.0, the disease is able to spread through the population. Larger values of *Ro* will cause a greater proportion of the Susceptible population to experience the disease and eventually end up in the Recovered population class. Although not shown in Figure 2, *Ro* also controls the speed at which the disease moves through the population. For many bioterror agents, (e.g., anthrax) the disease is not contagious at all and therefore *Ro* is zero.

The traditional SEIR equations as presented above have previously been adapted to include spatial components. Most commonly this adaptation has been made to include interactions between one or more sub-populations. These sub-populations can represent different cities within the same country (see Rohani et al., 1999; Bolker and Grenfell, 1996); however, the communication between the sub-populations is only made with a coupling parameter that is independent of the spatial location of the sub-populations or the distance between the sub-populations. Models that consider the interactions between discrete individuals have not been examined in detail and have generally focussed on simple nearest neighbor interactions (e.g., Market and Hare, 2000). The next section provides the basis for a SEIR model that describes the spatial and temporal evolution of the disease across a population of discrete individuals.

SPATIAL-TEMPORAL SEIR MODEL

The traditional implementations of the SEIR model, as discussed above, consider the population as a whole and do not consider the individual or the discrete nature of the links between individuals. As pointed out by Keeling and Grenfell (2000), the application of the traditional SEIR model assumes that each infectious individual interacts weakly with an infinite set of susceptible individuals. This assumption can be thought of as the population having an infinite uptake capacity for the disease. This assumption does not take into account any spatial structure in the susceptibility/immunity of the population nor the transmission efficiency of the disease. This section presents an adaptation to the traditional SEIR model to examine the spread of a contagious disease across a population of discrete individuals.

The progression of the disease and the state of the disease within any individual is modeled on a two-dimensional grid where each grid cell represents an individual and the entire grid represents the entire population being considered. In future applications, this grid of individuals could be mapped directly onto a spatial map of a city or region to directly link individuals with addresses. This type of mapping would also render some cells inactive to represent regions without any population (major interchanges, forests, etc.). This direct mapping of individuals with geographic information is not included in this report and is left to future applications. The dense, fully populated grid used in the following discussion is representative of a densely populated urban area. The grid-based approach will allow for relatively simple implementation of spatially variable susceptibility and immunity as well as providing concise control of the transmission of a disease between proximal individuals.

The SEIR equations presented above need to be reformulated in order to operate correctly on a population of discrete individuals rather than on a large population of perfectly mixed individuals. Instead of employing a global contact parameter, β, that describes the rate at which the disease spreads through the entire population, the spread of the disease must now be modeled as a series of interactions between individuals that are proximal to one another in the two dimensional space. In order to model this contact between two discrete individuals, the concept of an exposure likelihood function, *EL*, is presented.

The likelihood of a susceptible individual being exposed to a disease during a single time step is a function of the proximity of the susceptible individual to one or more infectious individuals. The likelihood of being exposed is also dependent on the range of personal interactions per individual within the population (essentially how many infectious individuals are in contact with the susceptible individual per time step) and the efficiency with which the disease is transmitted. There are a number of options available for constructing the formulation of the likelihood function including nearest neighbor difference equations and variations of the diffusion equation. For this work a series of functions used to model spatial covariance within the field of spatial statistics are used to define the exposure likelihood functions. These functions provide an extremely flexible and computationally efficient means of defining the exposure likelihood of multiple individuals in a grid-based population.

The first exposure likelihood function is based on the spherical model of spatial covariance:

$$\lambda(h) = \begin{cases} C - C\left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right), & 0 \le h \le a \\ 0.0 \quad , & a \le h \end{cases}$$

where C is the maximum possible value of the likelihood at zero separation distance between individuals. For a fully efficient transmission in a single time step, C= 1.0. The a parameter is the range of the likelihood function and controls the radius of influence over which a single infectious individual can transmit the disease. The separation distance between two individuals is h. In the current application of the model, the grid is composed of individuals and therefore h defines the proximity of the nearest infectious individual to any susceptible individual. The spherical likelihood function and the two others described here are shown in Figure 3.

Two additional exposure likelihood functions are also defined for use in this study. These two functions are the exponential exposure likelihood function:

$$\lambda(h) = C - C(1 - e^{-\frac{3h}{a}})$$

And the Gaussian exposure likelihood function:

$$\lambda(h) = C - C(1 - e^{-3\left(\frac{h}{a}\right)^2})$$

The three transmission likelihood functions with a range, a parameter, of 10.0 are shown in Figure 3. Note that the spherical model reaches a value of zero likelihood at the specified range value of 10.0. The exponential and Gaussian models reach the zero likelihood value asymptotically. The range parameter in the exponential and Gaussian functions defines the distance at which the likelihood decreases to 5 percent of the initial likelihood. Additional information on the background and application of these functions can be found in references dealing with geostatistics (e.g., Cressie, 1993; Deutsch and Journel, 1998; Olea, 1999) where they are the complement of what are known as *variogram* functions.

The likelihood functions defined above offer an extremely flexible means of modeling the spatial spread of a disease through a population. Because these functions are defined by analytical expressions, they are computationally very efficient relative to modeling the spread of the disease through the diffusion equation as has been done in other works. Additionally, they offer a wide range of flexibility in terms of the radius of influence and the efficiency of transmission from the infected individual to the susceptible individuals in the surrounding population. This flexibility is not limited to the simple functions shown above, but can be extended to any linear combination of the above functions with positive coefficients (see Olea, 1999, p. 80). It is also noted that the range of the likelihood functions need not be directionally isotropic. If, for example, a pathogen is transmitted along the direction of the prevailing wind, the ranges of the likelihood functions can be modified to honor this preferred orientation of transmission.

As presented thus far, the exposure likelihood functions are written to be solely a function of the separation distance between the infectious individual and any susceptible individual. These functions can also be adapted to be a function of the spatial location within the population domain as well as a function of time (e.g., the disease may not transmit with the same efficiency through the entire infectious period). These adaptations will be considered in future work.

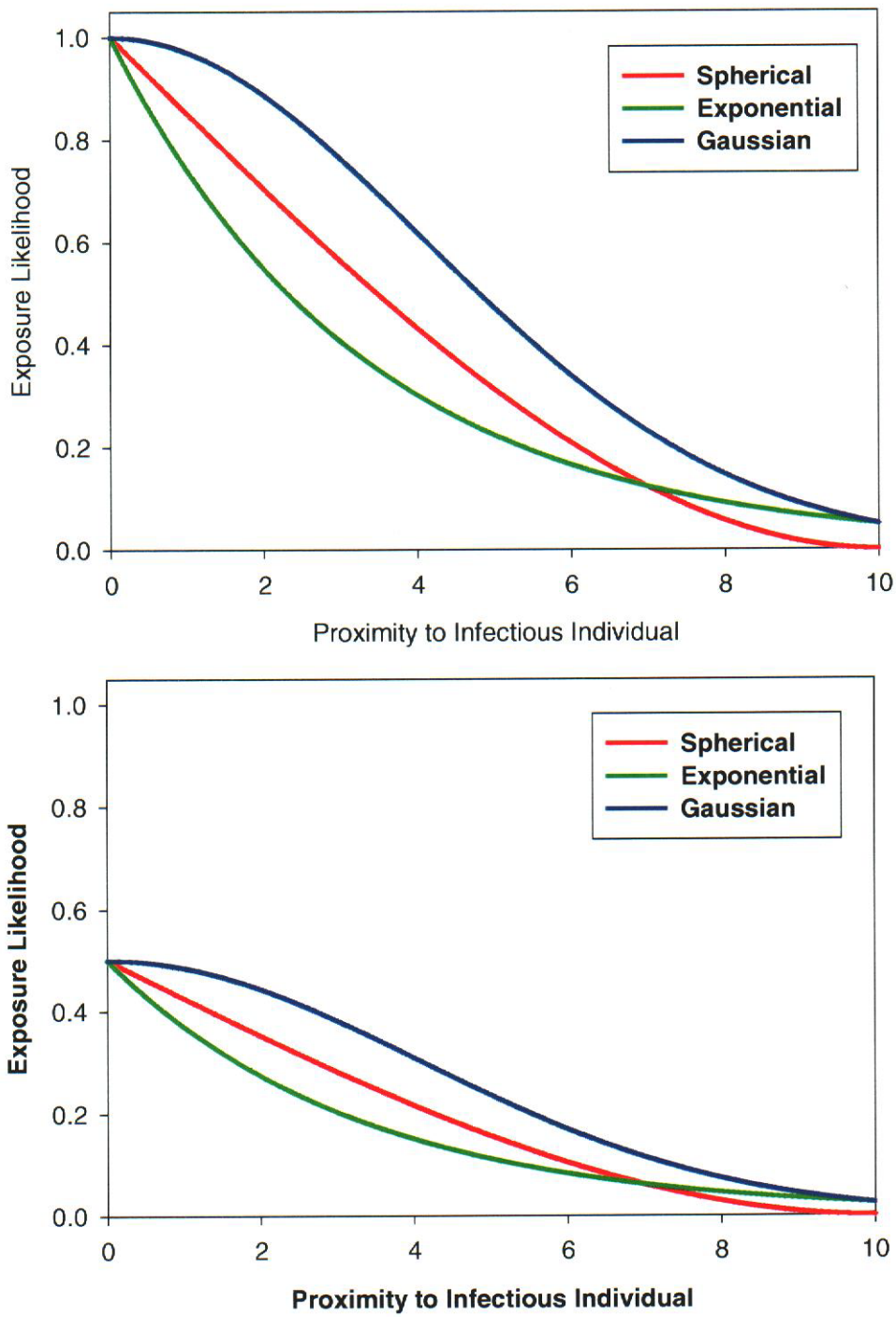

Figure 3. Comparison of the three exposure likelihood functions. The upper image demonstrates full disease transmission efficiency from the infectious individual. The lower image shows the functions for an individual with only 50 percent efficiency in disease transmission.

If a susceptible individual, i, is close enough to an infectious individual such that the likelihood of exposure reaches, or exceeds, 1.0 for that susceptible individual, then that individual moves out of the Susceptible and into the Exposed category:

$$\begin{bmatrix} if \sum_{t=1}^{t=1/\gamma} EL(i) \geq 1, & \Delta E = 1; & \Delta S = -1 \\ else & \Delta E = 0; & \Delta S = 0 \end{bmatrix}$$

Because the EL function is defined per time step, the EL function for individual i must be summed over the entire infectious period ($t = 1$ to $t = 1/\gamma$) to determine the likelihood of exposure.

The SEIR equations must be rewritten to describe the progress of the disease at the individual level. For each stage, the likelihood of any individual moving to the next stage begins at zero and is then incrementally increased as a function of exposure likelihood function, the incubation period, or the latency period.

For example, the likelihood of individual i moving from the susceptible to the exposed stage at time step j is termed $S \rightarrow E_{ij}$. This likelihood is the sum of the likelihood at time step $j-1$ and the value of the exposure likelihood function based on the proximity of any infectious individual(s). This likelihood is:

$$S \rightarrow E_{i,j} = S \rightarrow E_{i,j-1} + EL_{i,j}$$

Note that the EL_{ij} function takes into account all infectious individuals within the range a of the current susceptible individual. There is no need to account for the proportion of infectious individuals within the whole population as is done in the traditional SEIR model.

Similarly the likelihoods for making the other two transitions are:

$$E \rightarrow I_{i,j} = E \rightarrow I_{i,j-1} + \frac{\Delta t}{\sigma}$$

and

$$I \rightarrow R_{i,j} = I \rightarrow R_{i,j-1} + \frac{\Delta t}{\gamma}$$

Relationship Between Spatial Transmission and Ro

As discussed above, the reproductive ratio, *Ro*, is a global parameter that defines the temporal evolution of a disease in a population. This parameter is defined independently of any spatial structure within the population that may affect the transmission of the disease (e.g., a spatially variable susceptibility as a result of a spatially variable age-distribution). In this section, a linkage between the transmission likelihood functions and *Ro* is developed and tested.

The definition of the reproductive ratio is "the average number of secondary cases caused by an infectious individual in a completely susceptible population" (Anderson and May, 1992). This is an input parameter to the traditional formulation of the SEIR model. This definition of *Ro* implicitly assumes (after Keeling and Grenfell, 2000) that the total number of infectious cases is small relative to the number of susceptible individuals and that the distribution of infectious and susceptible individuals is uncorrelated.

In the spatially discrete model developed here, the theoretical value of *Ro* can be calculated as the total number of susceptible individuals within the range of influence, a, of a single infectious individual that become exposed over the infectious period. This is a theoretical *Ro* value in that it assumes all individuals within the distance a are equally susceptible and that each susceptible individual is only influenced by a single infectious individual. Obviously spatial variability in the susceptibility of the population, the existence of nearby individuals in stages of the disease other than susceptible and multiple infectious individuals within a distance of $2a$ will cause the actual value of *Ro* to deviate from the theoretical value. For further discussion of the assumptions inherent in the definition of *Ro*, see Keeling and Grenfell (2000).

In order to gain some perspective on the relationship of the theoretical value of *Ro* as a function of both the range of influence and the disease transmission efficiency a series of calculations were done. The theoretical value of *Ro* is calculated using various values of the range of influence from 1.0 to 5.0 at three different transmission efficiencies (1.00, 0.75 and 0.50) with a spherical *EL* function. The results of these calculations are shown in Figure 4.

The results in Figure 4 show that across all three values of transmission efficiency, the theoretical value of *Ro* can be roughly approximated by the square of the range of influence for range values up to 4 individuals. Figure 4 shows that in this discrete, deterministic, spatial model with an isotropic (circular) range of influence, the theoretical value of *Ro* can only be integer multiples of 4. Notably absent from the theoretical values of *Ro* are 16 and 32. This absence is due to geometric constraints due to the use of a circular *EL* on a square grid. Also from Figure 4, the level of transmission efficiency can significantly change the resulting value of *Ro* for the same range value. As an example, at a range of 3.5, the theoretical value of Ro is only 4 with a transmission efficiency of 0.50 but rises to 20 for transmission efficiencies of both 0.75 and 1.00.

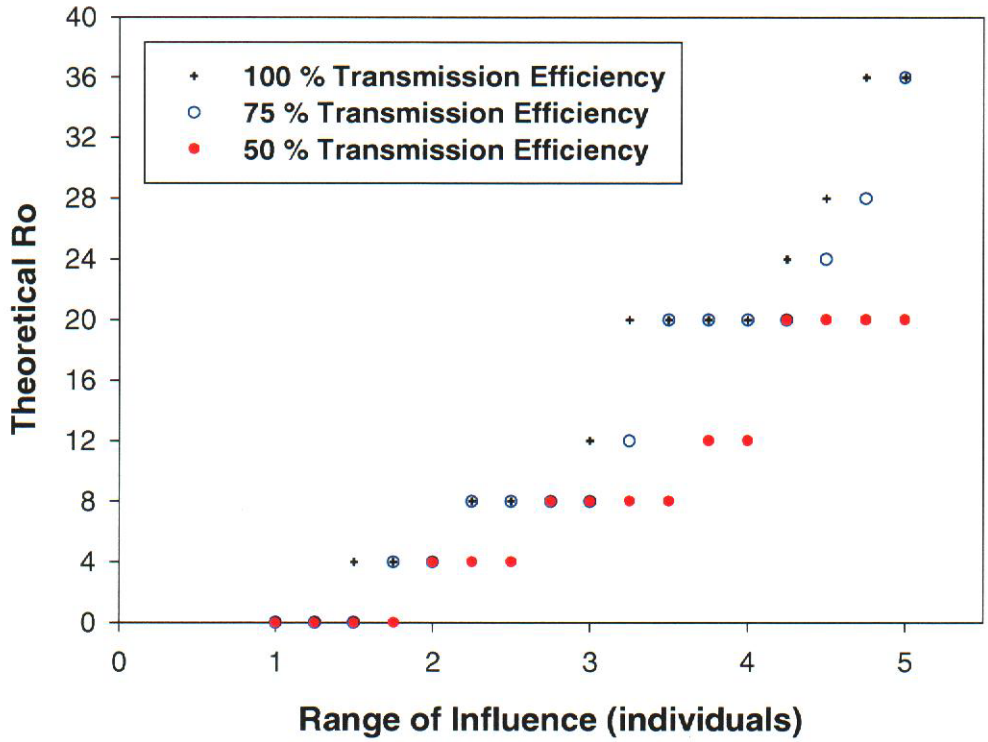

Figure 4. Theoretical values of *Ro* for different ranges of influence at three different disease transmission efficiencies. These results are for a spherical *EL* function.

Simulation Results

The spatial-temporal simulator is run to develop population curves and images of the spatial-temporal evolution of a disease through a population. The transmission efficiency is set to 1.0 (100 percent) and two ranges of influence, 2.0 and 3.0 individuals, are used to produce theoretical *Ro* values of 4.0 and 8.0 respectively. For both simulations, the population size is 40,000 and the initial fraction of the population that is exposed is 0.0025 (i.e., 100 people). These initial, exposed individuals are randomly distributed throughout the population. The incubation and infectious periods are 5.0 and 8.0 days respectively. The birth and death rates are set to zero and the time step is 1.0 days with a total simulation time of 250 days. With the exception of the initial exposed individuals, the entire population is susceptible with an initial likelihood of being exposed equal to zero. The epidemic curves resulting from these two simulations are shown in Figure 5. The epidemic curves show the sum of the exposed and infectious proportions of the population as a function of time.

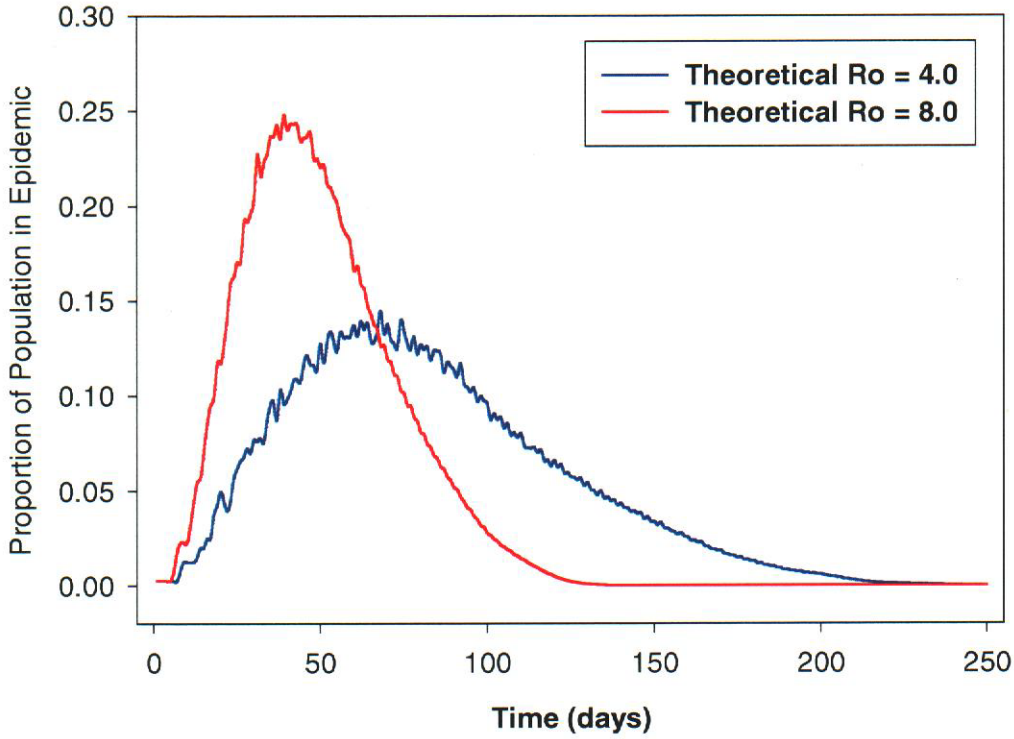

Figure 5. Epidemic curves created with the discrete spatial-temporal SEIR simulator for two different theoretical values of *Ro*.

The two different ranges of influence for the *EL* function create epidemics that move through the population at considerably different rates. The peak of the epidemic for the simulation with the theoretical value of *Ro* set to 4.0 is 74 days while the peak of the epidemic occurs at 42 days when the simulation is run with a theoretical value of *Ro* equal to 8.0. The peak proportion of the population within the epidemic is also higher (0.243 to 0.140) for the simulation with theoretical *Ro* set to 8.0 relative to the simulation with the theoretical peak set to 4.0. The spatial distribution of the stages of the disease at the peak of the epidemic are shown in Figure 6 for both simulations.

A single time slice of the disease evolution, as shown in Figure 6, is not the ideal way to view the spatial and temporal progress of the disease; however, several interesting results can be seen in Figure 6. At the initial time step, 100 individuals located randomly throughout the grid are exposed. The disease spreads out from these 100 individuals in a diffusion-like manner (Figure 6). This concentric spreading of the disease outward from the initially exposed individuals is due to the spatially homogeneous susceptibility of the population and the constant and deterministic specification of the SEIR parameters. In areas where multiple initially exposed individuals were located in close proximity, the

spread of the disease is faster. In these areas, the local value of *Ro* is higher than the theoretical value due to the additive effects of multiple *EL* functions.

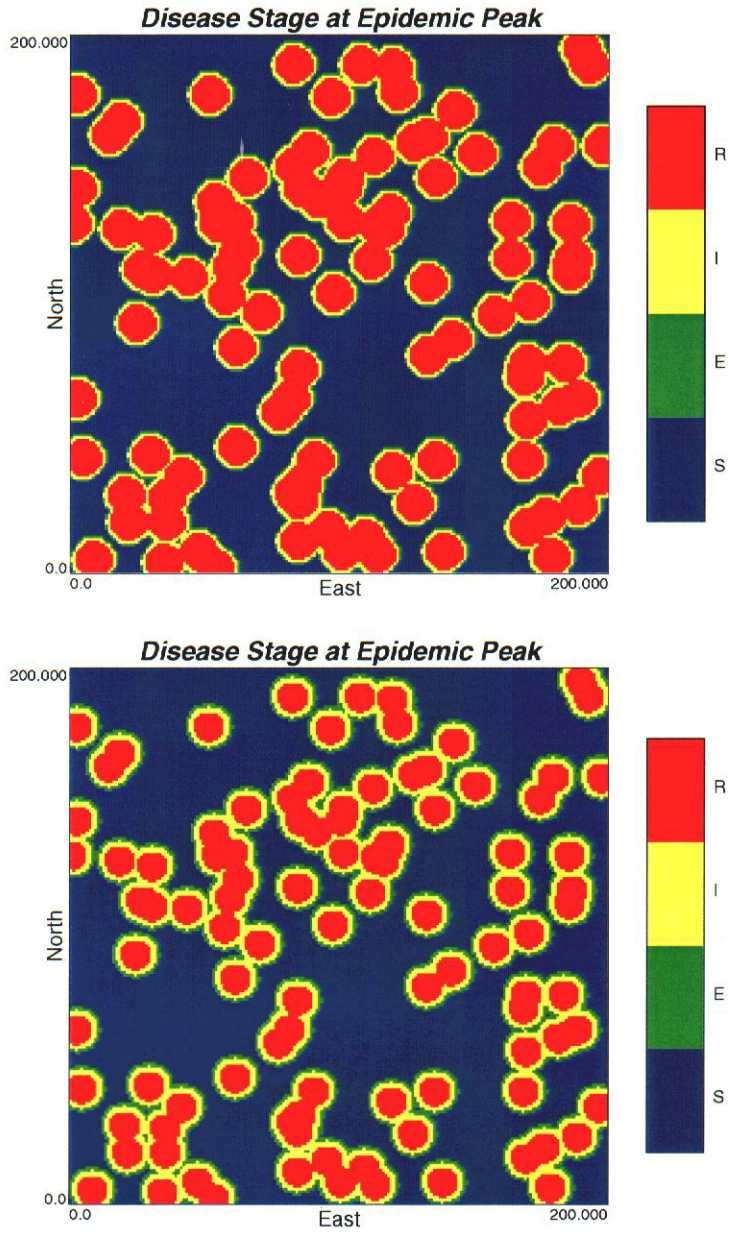

Figure 6. Spatial distribution of the disease stages at the time of the peak of the epidemic for simulations run with theoretical values of Ro equal to 4.0 (upper image) and 8.0 (lower image).

For the results in the upper image of Figure 6, the range of influence was specified as 2.0 and the theoretical value of *Ro* is 4.0. These parameters create a relatively slower moving epidemic and a short transition distance between the susceptible portions of the

population and the recovered portions of the population. The results in the lower image of Figure 6 were created with the range of influence set to 3.0 giving a theoretical *Ro* value of 8.0. This epidemic moves through the population much faster than the one simulated with a theoretical *Ro* of 4.0. The lower image of Figure 6 indicates that the larger range of influence spreads the epidemic faster and creates a wider transition zone between the susceptible and recovered members of the population. This wider transition zone is the members of the population that are either exposed or infectious (the epidemic). The images in Figure 6 demonstrate that the larger the range of influence, the larger the proportion of the population within the epidemic at any one time.

The results of the discrete spatial-temporal SEIR simulator developed here will be used to develop individual disease time histories for every individual in the population. One example of these individual disease time histories is shown in Figure 7 for a randomly located individual at grid cell location X = 50, Y = 50. The time history in Figure 7 was created in the simulation above with a theoretical *Ro* of 4.0. Figure 7 shows the exposure likelihood begin to increase from zero (Stage = 1) at about 47 days. This is due to the presence of an infectious individual within the range of influence of the example individual located at (50,50). The slope of this increasing likelihood changes through time as infectious individuals are within a of the individual located at (50,50). The exposure likelihood reaches 1.0 at day 52 and the disease stage changes to 2 (Exposed). From this point on, the slope of the line (the increase of the disease stage towards the next stage) is linear. This slope is $1/\sigma$ for the exposed stage (stage 2 to 3 in Figure 7) and $1/\gamma$ during the infectious stage (stage 3 to 4 in Figure 7).

The disease time history as shown in Figure 7 is available for all individuals in the population. Probabilistic estimates of the different syndromes, symptoms, and signs that an individual would develop at each stage of a specific disease can be derived from these time histories. These estimated syndromes, symptoms, and signs will comprise the S^3 data that will go into the pattern recognition portion of this project. Techniques for determination of the probabilistic estimates of the S^3 data will be completed in the future and be linked to the time histories produced by the discrete spatial-temporal SEIR simulator.

One attribute that may prove useful in discriminating one disease from another in the pattern recognition phase of this project is the temporal derivative across the population domain. For the finite time stepping scheme used here, the temporal derivative for individual I is simply:

$$deriv(i) = \frac{\left| stage_j - stage_{j-1} \right|}{\Delta t}$$

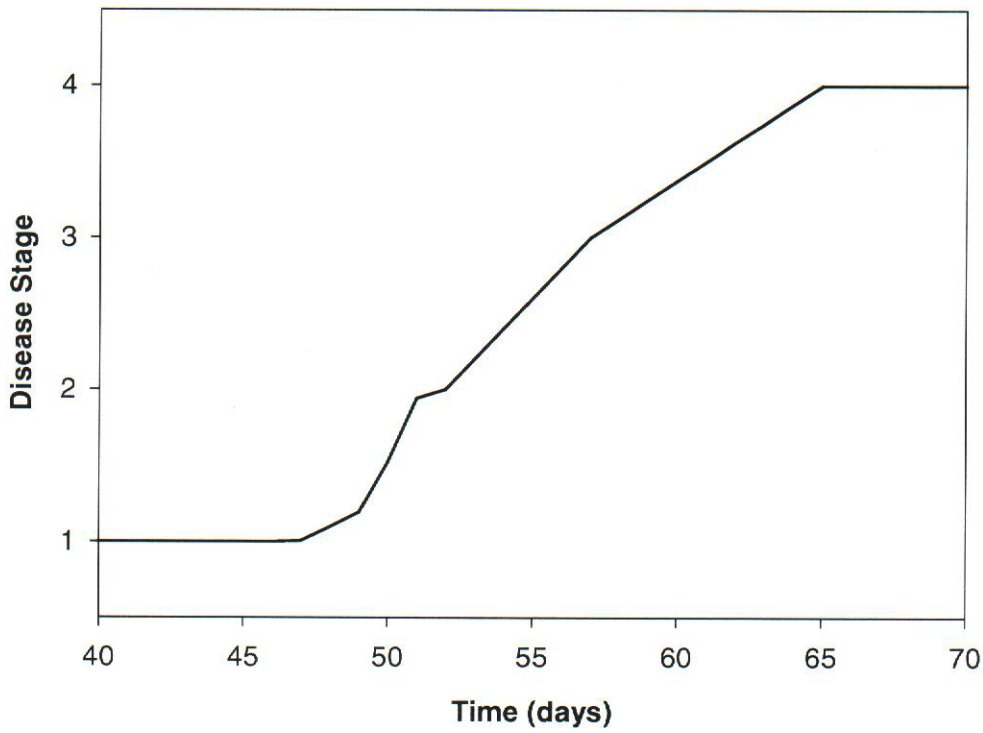

Figure 7. Disease progression time history for an example individual located at (50,50). The numbers on the Y-axis define the disease stages: 1= S, 2 = E, 3 = I, 4 = R.

where $stage_j$ and $stage_{j-1}$ are the levels of the disease stage (see Figure 7) at time steps j and $j-1$ and Δt is the length of the time step. The differences in the temporal derivative for the time step corresponding to the peak of the epidemic are shown for the simulations created with theoretical values of Ro equal to 4.0 and 8.0 in Figure 8. In general, the larger value of Ro produces lower values of the derivative and a wider band of non-zero values.

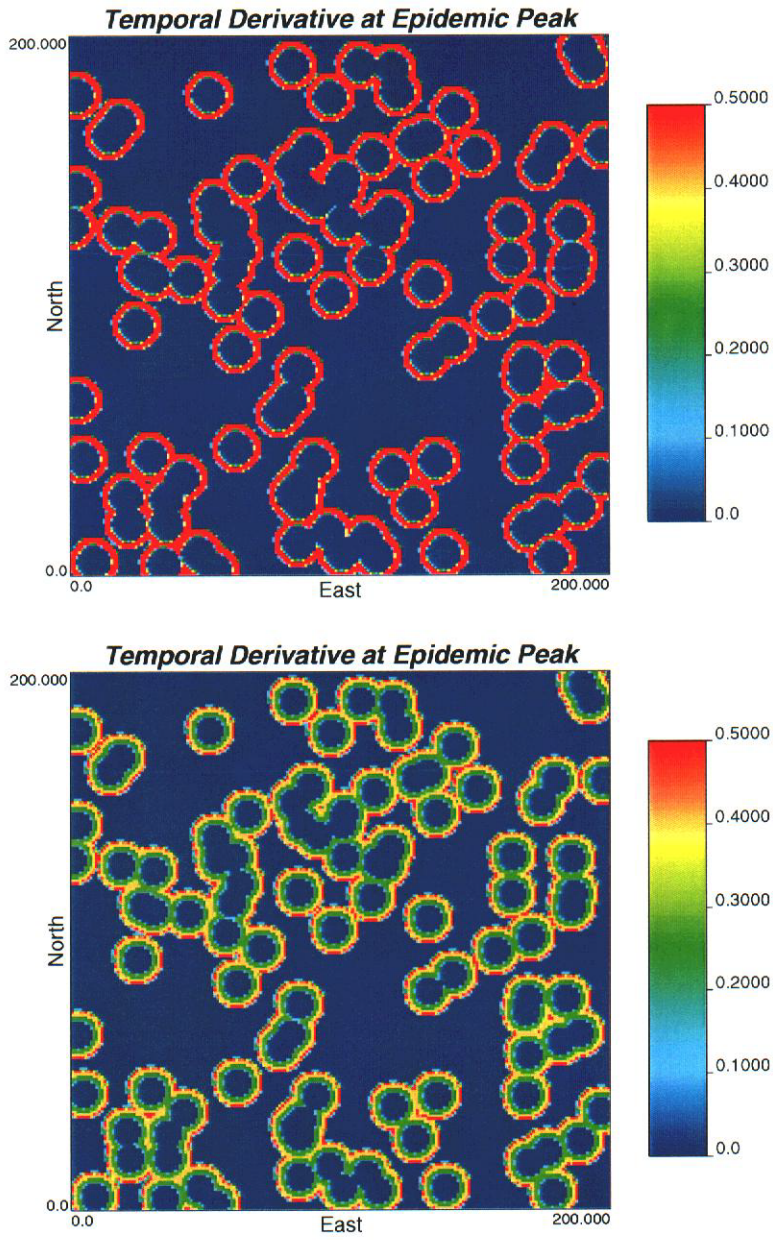

Figure 8. Temporal derivative values across the population domain at the time step corresponding to the peak of the epidemic for the simulations with the theoretical value of *Ro* equal to 4.0 (upper image) and 8.0 (lower image).

Spatial Variability

One attractive advantage to developing a spatial temporal simulator that works on a discrete grid of individuals is that environmental information that is available over the same spatial domain and that may effect the spread of the disease can be readily incorporated into the model. A simple example of this is demonstrated below with information on the spatial variability of individual susceptibility across the domain. This spatial variation in susceptibility may be due to variation in the ages of the individuals or other factors.

Spatial variation in the susceptibility of the individuals is simulated by drawing uniformly distributed random numbers between –1 and +1 and assigning these random numbers as the likelihood of being exposed. In the simulations discussed previously in this report, the initial exposure likelihood of all individuals was set to zero across the entire domain (with the exception of the small fraction of initially exposed individuals). By assigning likelihood of exposure values between –1 and +1 through a random number generator, the individual resistance to exposure can be much greater (-1) or essentially zero (1.0). In this example, the spatial variability in the initial susceptibility of the population is uncorrelated from one individual to the next.

Two different simulations are run with the same pattern of spatial variation in the initial level of susceptibility. The parameters for both simulations are the same as those in the previous section with the spatially uniform initial susceptibility with the exception of the parameters controlling the exposure likelihood function. For both of these simulations with spatial variability in the degree of susceptibility, the efficiency of the disease transmission is set to 0.20 (in the above simulations it was 1.00). In the first simulation, the range of the *EL* is set to 2.0, which, along with a transmission efficiency of 0.20, defines a theoretical *Ro* of 0.0. This value of *Ro* would not be high enough to propagate the disease beyond the initially exposed cases if the initial susceptibility was set to zero at all other cells. In the second simulation, the range is increased to 3.0 but the transmission efficiency of only 20 percent keeps the theoretical value of *Ro* at zero. The epidemic curves for these two simulations are shown in Figure 9.

As seen in Figure 9, the effect of adding spatial variability to the initial values of susceptibility across the grid has an extreme effect on the propagation of the disease through the population. For both simulations the theoretical value of *Ro* is zero and with a spatially uniform initial susceptibility, there would be no propagation of the disease. The addition of spatial variation in the initial levels of susceptibility allows the disease to propagate through the population.

The change in the range of influence from 2.0 to 3.0 also has a drastic effect on the epidemic curve. The range of 2.0 creates an epidemic that propagates extremely slowly through the population. The peak proportion of the population experiencing the epidemic is less than 2 percent (peak occurs at 270 days). Although not shown in Figure 9, at 1500 days, the proportion of the population that is susceptible is still near 10 percent. This low

range value coupled with spatial variation in the initial susceptibility level creates a disease that reaches some sort of equilibrium and can be described as endemic. The increase in the range to 3.0 creates a much faster moving epidemic that peaks at 61 days and moves through the entire population in less than 250 days. Keep in mind that the theoretical *Ro* value equals zero for both simulations.

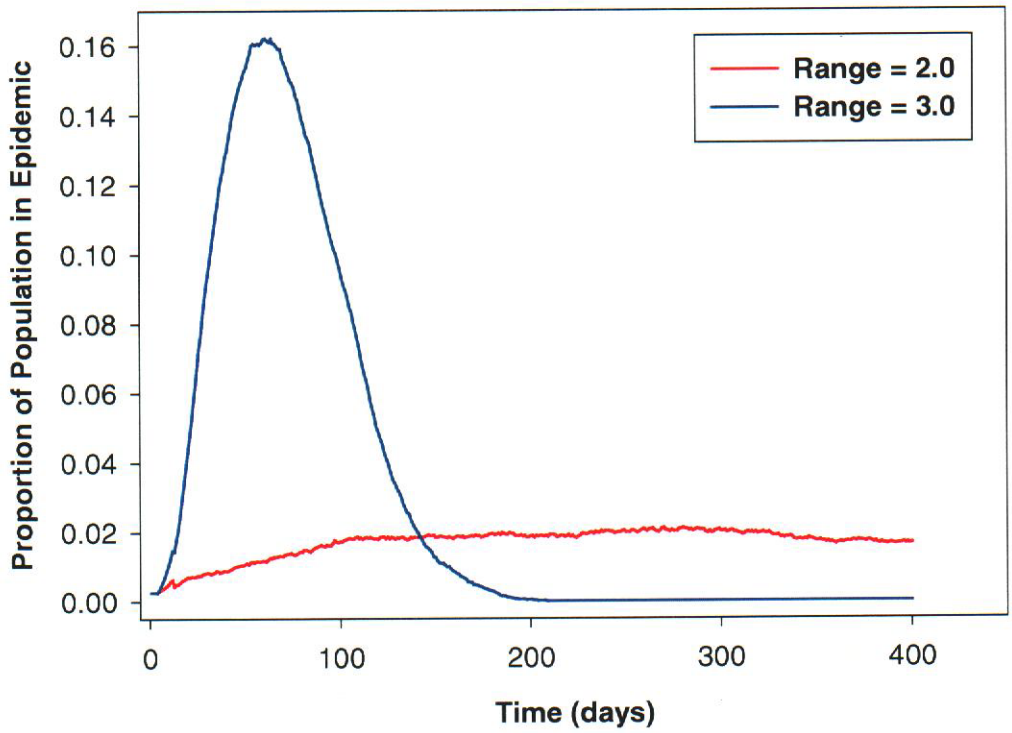

Figure 9. Epidemic curves for two simulations done with the same spatial variation in the initial susceptibility values but different ranges of influence. Note that both simulations have a theoretical *Ro* of 0.00.

The spatial maps of the disease stage at the epidemic peaks are given for both simulations in Figure 10. The differences in the disease stage images at the epidemic peaks are significant. The image from the simulation created with a range of influence in the exposure likelihood function of 2.0 (upper image, Figure 10) shows an extremely irregular pattern in the distribution of exposed, infectious and recovered individuals. The results from a range of 3.0 (lower image, Figure 10) exhibit a more regular spreading pattern away from the initially exposed individuals.

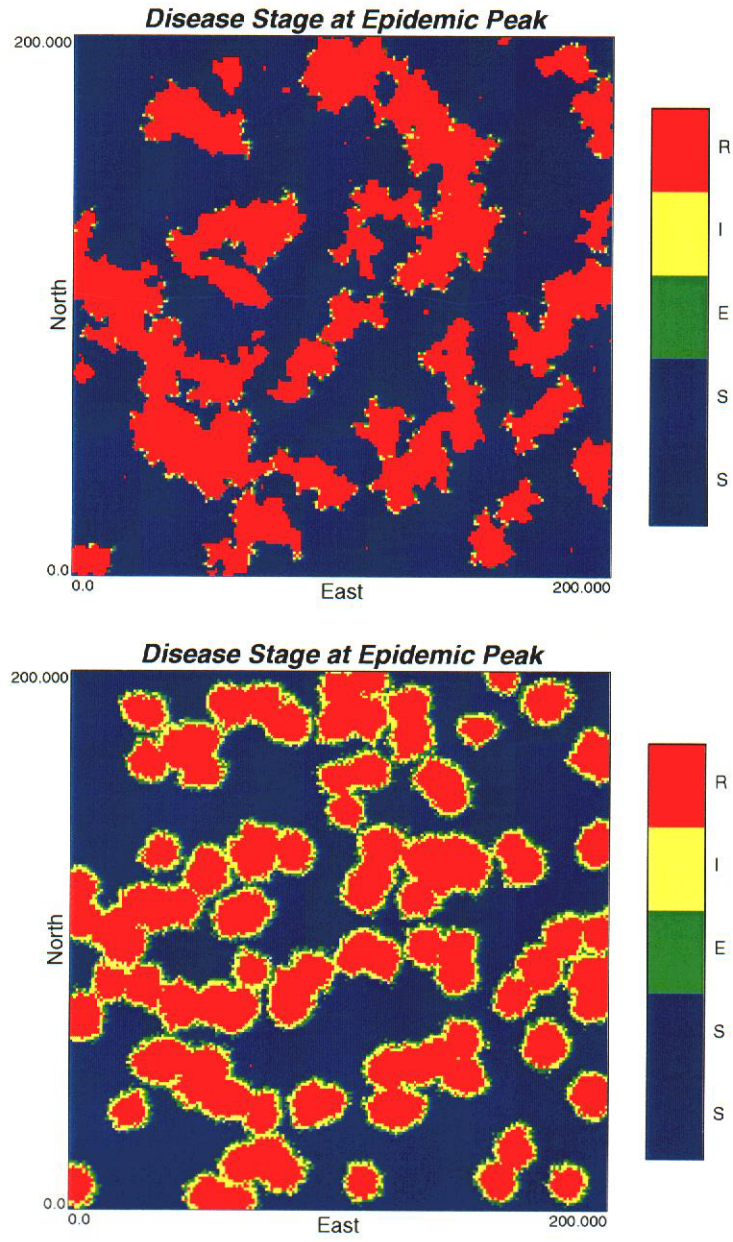

Figure 10. Disease stage images at the epidemic peak for simulations with spatial variation in the initial levels of susceptibility. The upper image results from an exposure likelihood function with a range of 2.0 and the lower image results from an exposure likelihood function with a range of 3.0.

The results created with a range of 2.0 (upper image, Figure 10) show a number of isolated recovered individuals. These are individuals that were initially exposed, but were unable to transmit the disease to any surrounding individuals due to these surrounding individuals having a relatively low susceptibility (high immunity) to the disease. These individuals may be difficult to discriminate from individuals infected during a bioterror attack with a non-contagious agent such as anthrax that is not capable of spreading between individuals.

The significant differences in the epidemic curves shown in Figure 9 indicate the profound effects that variability in the initial susceptibility of the population can have on the spread of a disease. The epidemic curve resulting from a range of 2.0 indicates an endemic expression of the disease. The fact that only a small proportion of the population is infected at any given time would make it very difficult for a disease surveillance network to readily identify this disease.

Summary

This report has summarized the progress made to date on the development of a discrete spatial-temporal SEIR simulator. The concept of the SEIR model that has been applied to populations, or somewhat loosely coupled sub-populations, has been extended to work on a population of discrete individuals located on a two-dimensional grid. The propagation of the disease across this grid and through time is controlled by an exposure likelihood function developed for this project from spatial covariance functions used in other fields. The exposure likelihood function controls the spread of the disease from one individual to surrounding individuals. The connection between the exposure likelihood function and the reproductive ratio, *Ro*, used in the traditional SEIR model has been developed. A series of numerical experiments using the discrete spatial-temporal SEIR simulator was completed to determine the theoretical value of *Ro* corresponding to different parameterizations of the exposure likelihood function. The current version of the discrete spatial-temporal SEIR simulator is capable of producing time histories of the progression of the disease for every discrete individual. These time histories will be used as the basis for assigning the proper S^3 data to each individual. In order to provide the S^3 data, the simulator will be linked to a probabilistic S^3 estimation technique. The addition of spatial variation in the initial levels of susceptibility can drastically alter the spread of the disease. Spatial variation in the susceptibility pattern can also allow for the spread of the disease even when the theoretical *Ro* value is zero.

Future Work

Several additional components of the discrete spatial-temporal SEIR simulator could be added. The first of these is to determine the SEIR parameters (e.g., σ, γ, *Ro*) for a number of common diseases as well as diseases that might be the introduced in a terrorist attack. These parameters will be used to develop the spatial-temporal histories of these

diseases across a discrete population. These histories will be used to generate the S^3 data as input to the pattern recognition portion of the project.

It is necessary to keep in mind that any individual patient will probably only receive medical attention once for the treatment of a more common disease. For a fatal disease, such as that introduced through a bioterror attack, the patient's association with the medical community will obviously be more frequent as the symptoms progress. The correct amount of decimation in time of the complete S^3 data set produced by the discrete spatial-temporal SEIR simulator must be determined.

The spatial coordinates of each individual in this model are provided at a finer resolution than will be available through the actual collection of the S^3 data that is limited to the resolution of individual zip codes. In order to simulate this level of spatial aggregation, data from the discrete individuals will be accumulated within polygons of arbitrary shape and size representing zip codes. This technique could also be used to accumulate data at the level of counties, school districts or other political boundaries.

Additional developments of the discrete spatial-temporal SEIR simulator will focus on adding the capability to examine spatially variable levels of both susceptibility and immunity to certain diseases. This will be done with fairly simplistic random fields as well as with spatially heterogeneous fields created through spatial estimation algorithms borrowed from the field of geostatistics. Several authors have proposed the concept of spatially variable age structure controlling the spatial variability of susceptibility. Information on the spatial distribution of ages within a major metropolitan area(s) could be obtained and used in creating the spatially variable susceptibility maps for use in the discrete spatial-temporal SEIR simulator.

In its current state, the discrete spatial-temporal SEIR simulator is capable of simulating the spatial and temporal evolution of a single disease across a population. It will be necessary to simultaneously model different diseases in order to determine if the proposed pattern recognition techniques can discriminate the outbreak of an anomalous, terrorist-induced disease against a background of regular diseases such as influenza. Using the current configuration of the simulator, it is possible to model the simultaneous spread of independent diseases by running the simulator multiple times (once for each different disease). If the pathogen introduced by a terrorist organization operates independently of any common background diseases, this approach will work. However, if the propagation of a terrorist induced pathogen is correlated with the stages of a common disease, then an extension to the simulator will be necessary. For example, individuals who are suffering from the flu (in the exposed or infectious stage) may have a much higher susceptibility to contracting another infectious disease including an infection delivered by a terrorist organization.

In terms of the final use of spatial temporal modeling in this project, the patterns detected through a pattern recognition procedure will be used as the basis for spatial and temporal *prediction* of the spread of a disease. Instead of using the SEIR model as written here, it will be adapted to incorporate real-time information into a predictive model. The

incorporation of real time data will lead to a more data driven approach to modeling. Ity is proposed that this predictive model will use some combination of geostatistical simulation algorithms to incorporate the real time data and the SEIR model to incorporate the processes governing the spread of the disease.

References:

Anderson, R.M. and R.M. May, 1992, *Infectious Diseases of Humans*, Oxford Science Publications, Oxford.

Bartlett. M.S., 1957, Measles periodicity and community size, *Journal of the Royal Statistical Society, A*, Vol. 120, pp. 48-70.

Bolker, B.M. and B.T. Grenfell, 1996, Impact of vaccination on the spatial correlation and persistence of measles dynamics, *Proceedings of the National Academy of Sciences, USA*, Vol 93, pp. 12648-12653.

Cressie, N.A.C. 1993, *Statistics for Spatial Data, Revised Edition*, John Wiley and Sons, Inc., New York, 900 pp.

Deutsch, C.V. and A.G. Journel, 1998, *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 380 pp.

Earn, D.J.D., P. Rohani and B.M. Bolker, 1999, A simple model for complex dynamical transitions in epidemics, *Science*, (287) 667.

Grenfell, B.T., 1992, Chance and chaos in measles epidemics, *Journal of the Royal Statistical Society, B*, Vol 54, No. 2, pp. 383-398.

Hamer, W.H., 1906, Epidemic disease in England, *Lancet*, i, pp. 733-739.

Keeling, M.J. and B.T. Grenfell, 2000, Individual-based perspectives on Ro, *Journal of Theoretical Biology*, Vol. 203, pp. 51-61.

Olea, R., 1999, *Geostatistics for Engineers and Earth Scientists*, Kluwer Academic Publishers, Holland, 328 pp.

Rohani, P., D.J.D. Earn and B.T. Grenfell, 1999, Opposite patterns of synchrony in sympatric disease metapopulations, *Science*, (286) 968.

Soper, H.E., 1929, Interpretation of periodicity in disease prevalence, *Journal of the Royal Statistical Society*, Col. 92, pp. 34-73.

Distribution List for SAND2000-2232

Organization 5353 MS 1203
Al Zelicoff

Oranization 5907 MS 1219
Annette Sobel

Organization 6100 MS 0701
Wendy Cieslak
Peter Davies

Organization 6115, MS 0735
Sean McKenna (15 copies)
Erik Webb

Organization 6805, MS 0716
Chris Olson

Organization 6851, MS 0778
George Barr

Organization 8000 MS 9001
Greg Thomas

Organization 8114 MS 9201
Dawn Kataoka
Todd West

Organization 15334 MS 1188
John Wagner

Organization 15352 MS 0843
Drayton Boozer
Kevin Green
Mark Koch (3 copies)

Organization 16000 MS 0839
Tom Karas

Technical Library: Organization 9616 MS 0899 (2 copies)

Central Technical Files: Organization 8945-1 MS 9018

Review and Approval Desk: Organization 9612 MS 0612
For DOE/OSTI